《科学学研究》已录用，待刊发表

“睡美人”文献的重要特征、预测线索与政策启示*

杜建[1]，武夷山[2]

（1. 中国医学科学院医学信息研究所 北京 100005；2. 中国科学技术发展战略研究院 北京 100038）

摘要：睡美人文献研究对于图书馆界文献战略保藏、科技界早期识别变革性研究与缩短重大科学发现的认可时滞具有重要意义。通过数据分析和案例研究揭示了睡美人文献具有多出自跨学科研究和综合性期刊、多具有潜在技术与应用属性、多为高质量研究三大特征。初步凝练出预测睡美人文献的若干关键线索：一是识别变革性研究并追踪其技术转化应用状况，包括监测作者是否持续开展该主题的研究，是否从理论研究拓展到实践研究，实践中是否成功等；论文发表之后是否有专利授权，论文是否被专利引用等。二是笔者提出的一个用于识别非高被引论文中睡美人文献的无参数指标——Bcp 指数，能够识别出那些正处于“沉睡-唤醒”萌芽期的论文，特别是长期沉睡后初现被引突增苗头，且总被引次数尚未成规模的论文。基于睡美人文献重要特征与预测线索的分析，从加强变革性研究、识别研究前沿和改善学术评价三个方面讨论了睡美人文献研究的政策启示。

关键词：睡美人文献；变革性研究；专利引用；预测线索；研究前沿；学术评价

Sleeping Beauties in Science: Key characteristics, early identification clues and science policy implications

Du Jian[1] Wu Yishan[2]

[1] *Institute of Medical information & Library, Chinese Academy of Medical Sciences, Beijing(China)*

[2] *Chinese Academy of Science and Technology for Development, Beijing(China)*

Abstract: The study on systematic identification of “Sleeping Beauty” (SB) publications and on their awaking mechanisms will promote the strategic preservation of documents in libraries, and provide implications for identifying potential “ahead of time” discoveries or transformative research, and shortening time lag for original research to get recognized. Based on the data analysis and case studies, this paper figured out some key characteristics of SB publications. The content of SB papers tends to be multi-discipline research and most of which tend to be published in prestigious multidisciplinary journals with higher impact and wider scope. One is perhaps more

---

* 国家自然科学基金项目“睡美人文献及唤醒睡美人的王子文献的识别方法与典型特征研究”（项目编号：71603280）

作者简介：杜建（ORCID: 0000-0002-7621-9995），男，1986 年生，博士，助理研究员，研究方向为科学计量学与医学科研评价；武夷山（ORCID: 0000-0001-6271-7881），男，1958 年生，研究员，博士生导师。研究方向为科学计量学与科技政策。

inclined to believe that SB relate to more fundamental and basic, and less to application-oriented work. But a surprising finding is that half of the SBs are application oriented and significantly more cited in patents than ‘normal’ papers, which demonstrated the potential technical and application-research properties of SB. The key publications and pioneering works of the recent Nobel Prize winners for physiology and medicine are delayed-cited rather early-cited in the entire citation life-time of an article. We further suggested several clues for early identifying SB. First, Transformative research tend to be neglected or resisted by the scientific community initially and this neglect or resistance could be regarded as the key clue for the early prediction of SB literature. Publications belonging to so-called transformative research, even when less frequently cited than others, should be given special attention as early as possible, because they may suddenly attract many citations after a period of sleep. One could identify transformative research through some text terms (such as "disagree", "overcome", "break", "dispute"...). In order to discern such potential transformative research, we could observe whether the relevant documents get early citation from patents or not, because many SB documents tend to be technical research in nature and application-oriented. Second, the Bcp index proposed in our foregoing paper is more sensitive in identifying the “lower level SBs”, which refers to the case when the total citations and the maximum annual citations of SBs are not so high in comparison with other typical SBs. This index may be used to identify the papers at the sleeping-awakening interface, which provided new tools for arousing attention of the science community to previously overlooked but important research. We also pondered some policy implications about the associations of SB publications with transformative research, research front and research evaluation.
Keywords： Sleeping Beauty publications; transformative research; patent citing; early identification clues; research front; research evaluation

“睡美人”（Sleeping Beauties）文献是指一篇文献在发表后的相当长一段时期内零被引或低被引，仿佛在沉睡，而后突然高被引，就像被唤醒一样。唤醒睡美人的文献被称为王子文献[1]。睡美人文献通过动态反映文献被引的时序特征和历史过程，从科学计量学角度对科学社会学领域延迟承认现象作了定量描述。科学和技术领域都有延迟承认现象，在科学中表现为睡美人文献，在技术中表现为睡眠专利（sleeping patents）[2]。文献引用符合“长尾分布”，睡美人文献本身是高被引文献，本研究既考虑了“头部”（高被引），也考虑了“长尾”（低被引和零被引）。针对睡美人文献被引生命周期全谱段的研究是对引文分析的深化，可以从新的角度反映论文的质量与影响。

从截至目前有关睡美人文献的大规模的分析来看，在严格的人为参数界定下，睡美人文献的比例约为万分之一到千分之一，而对于诺贝尔奖得主发表论文，睡美人文献的比例约为百分之一。笔者认为睡美人文献并非以往所认为是罕见的，情报学和科学计量学领域对高被引文献的界定也通常选取被引次数排名前 1%、前 1‰的文献，而睡美人文献也是这样的比例。如何能早期预测睡美人文献，找到睡美人文献的早期标志，提示图书馆界长期收藏，提示科学界及早关注，对于促进科学家潜心创新、规避延迟承认将具有重要意义，在实践中对于基金资助机构和科技决策者发现超前性或变革性创新，提前部署相关的前沿研究和科技规划也将有重要参考价值。

本文通过案例研究和数据分析总结睡美人文献的重要特征，试图能找到睡美人文献的若干预测线索，并由此讨论睡美人文献研究的科技政策启示。这样，在潜在睡美人文献被引次数很少时就能够及时发现其睡美人潜质并推荐给学术共同体，以缩短重要科学发现的认可时滞。

1 睡美人文献的重要特征研究

根据直观理解，人们往往认为睡美人文献多倾向基础性、前沿性和理论性的文献，提出的高深理论或概念超前于当时条件与认知水平，因此不被理解而遭遇延迟承认。因而睡美人文献似乎应该多为基础研究而非应用研究，似乎应多具有科学属性而非技术属性。但科学和技术是相互依存、相互促进的。科学研究是追求真理和知识创新的活动，并以其研究成果提升社会实践为目的。任何科学研究都需要真正的施惠人类社会才是其最终价值的体现[3]。笔者假设，既然睡美人文献总是与重大科学发现相关联，而这样的科学发现之所以重要是因为它必定具有实践的属性。因此，睡美人文献是否也应该具有技术与应用属性，而非纯科学属性？本文将在现有相关研究基础上，继续通过典型案例和小规模数据来验证这一假设。

1.1 多出自跨学科研究和综合性期刊

Ke et al. (2015)做了至今最大规模的睡美人文献研究，对 1900s 以来 Web of Science 收录的被引次数至少为 1 次的 2200 多万份发表在自然科学及社会科学领域的论文进行系统分析后发现[4]，物理学、化学和数学贡献了最多的睡美人文献，睡美人文献的研究内容多具有学科交叉的特征。产出睡美人文献最多期刊是《自然》《科学》和《美国科学院院刊》，均为综合性期刊。原因在于综合性期刊的多学科交融性，因此能在几十年后还受到持续的关注，等到某年在机缘巧合之下浮出水面。在这三种杂志上发表的高水平论文可能是过于超前的思想，无法被当时的学术界理解；或是偏向未来主义的大胆预测，直到预言证实之日才得以苏醒。

1.2 具有潜在技术与应用属性

1.2.1 多为应用导向型研究且常被专利引用

数据量位于第二的睡美人文献研究是 van Raan(2015)从 Web of Science 数据库 1980-1994 年发表的论文中（被引次数统计截至 2013 年，被引至少 20 年），按照一个比较严格的阈值——[10, 1, 10, 5]，即发表之后 10 年内年均被引次数不超过 1 次，唤醒后 10 年后年均被引次数要大于 5 次，来识别物理学、化学、工程学与计算机科学领域的睡美人文献。认为睡美人文献多是应用型（技术）创新。

按上述标准，共识别出物理学领域睡美人文献 389 篇，化学领域 265 篇，工程与计算机领域 367 篇。物理学领域睡美人文献中有 210 篇（占 53%）被划分到至少一个应用型研究领域，即超过一半的睡美人文献具有应用研究型倾向。基于荷兰莱顿大学提出的期刊分类系统，将期刊分为基础研究和应用研究，对这些睡美人文献的发表期刊分析发现，物理学领域睡美人文献，有约一半发表在应用型研究期刊上，化学领域睡美人文献中，70%发表在应用型研究期刊上；在工程和计算机领域，全部论文均发表在应用型期刊上。可见，物理学、化学、工程学和计算机科学领域的睡美人文献多为应用导向型研究[5]。Van Raan 将其称为潜在的睡美人创新（Sleeping Innovations）。采用 PATSTAT 数据库检索这些文献是否被专利引用过[6]，发现 3 个学科领域睡美人文献被专利引用的比例（16%，23%，29%）远高于所有文献的平均水平，在 Web of Science 数据库中，大约 4%的文献会被专利引用。可见睡美人文献被专利引用的次数要比“普通”论文被专利引用的次数更多，即睡美人文献更易被专利引用。

受 van Raan 以上研究的启发，笔者继续探讨，既然睡美人文献具有潜在的技术与应用属性，那么睡美人文献在发表之初的低被引或零被引阶段是否会被专利引用？被专利施引年是否早于唤醒年？

1.2.2 被专利首次引用年常早于唤醒年

（1）2014 年诺贝尔化学奖关键论文被专利引用情况

笔者前期相关工作已发现[7]，2014 年诺贝尔化学奖关键论文——Hell SW 于 1994 年发表于 Optical Letter 的论文是一篇典型的睡美人文献，简称 Hell(1994)。基于 Lens.org 专利-论文引用检索平台，截至 2017 年 2 月 14 日，Hell(1994)这篇文献共被 67 项专利、50 个专

利族所引用，其中首项施引的专利就是 Hell Stefan 和 Wichmann Jan 于 1995 年 2 月 1 日向世界知识产权组织申请的 PCT 专利，名称为 Process and device for optically measuring a point on a sample with high local resolution，1995 年 8 月 10 日公开。专利发明人和论文的作者完全相同。该专利于 1994 年 2 月 1 日最早申请了德国专利。同时于 1996 年 11 月 21 日向美国专利局申请，并于 1998 年 3 月 24 日获得美国授权（US 5731588）。1997 年 10 月 22 日申请了欧盟专利，并于 2001 年 8 月 8 日获得欧盟授权。

图 1 2014 年诺贝尔化学奖关键论文——Hell(1994)被专利引用情况

这项专利仅引用了 2 篇非专利文献[8,9]，而这两篇文献恰好就是获得诺贝尔奖的 2 篇关键文献，这两篇文献同时又是睡美人文献，其唤醒时间是 2000 年[10]。可见，该研究从一开始就具有科学-技术双重属性。

（2）Science 和 Nature 睡美人文献被专利引用情况

笔者前期提出了一个识别睡美人文献的新的无参数指标，并用该指标识别出 Science 和 Nature 上的 Top10 睡美人文献[11]，试图探索一下这些睡美人文献是否被专利引用过。通过 Lens.org 检索论文被专利引用的情况，PubMed 数据库已经提供了 Lens.org 中 PubMed 论文与专利引用的链接。基于同族专利统计被引次数,首次施引专利界定为首项授权专利（publication date）。表 1 可见，10 篇睡美人文献中，6 篇被专利引用过。其中被专利引用次数最多的是第一篇，受到 49 个专利族的引用。专利被引时滞是指专利优先权年（最早申请年）减去论文发表年。以上的案例分析表明，睡美人文献被专利施引的年份均早于论文引用下的唤醒年。

表 1 Science 和 Nature 上 10 篇睡美人文献被专利引用的情况

期刊来源	被引次数	唤醒年	专利族被引次数	首次施引专利申请年	专利族数	授权年	优先权年	优先权号	被引时滞
Nature, 1972, 238(5358), P37	10776	2004	49	Nov 17, 1975	9	Mar 8, 1977	Nov 17, 1975	US 63255775	3
Science, 1970, 170(3957), P497	255	2002	-						
Nature, 1973, 244(5414), P281	415	2005	6	Sep 26, 1983	8	Oct 7, 1986	Sep 27, 1982	GB 8227500	9

期刊来源	被引次数	唤醒年	专利族被引次数	首次施引专利申请年	专利族数	授权年	优先权年	优先权号	被引时滞
Science, 1971, 171(3977), P1212	553	2002	-						
Nature, 1970, 227(5258), P561	602	2003	31	May 3, 2002	1	Oct 10, 2006	Sep 10, 1998	US 9976398	28
Nature, 1970, 225(5232), P563	241	2005	2	Sep 30, 1994	2	Oct 20, 1998	Jul 17, 1986	US 88707086	16
Science, 1972, 177(4047), P393	888	2002	-						
Nature, 1970, 227(5257), P520	736	2003	-						
Nature, 1971, 229(5286), P560	286	1999	1	Sep 27, 2001	1	Jan 20, 2004	Sep 27, 2000	US 23604500	29
Science, 1970, 167(3916), P279	423	2004	6	Dec 27, 1988	15	Sep 4, 1990	Dec 27, 1988	US 29030788	18

注：专利被引次数截至 2017 年 2 月 14 日，数据来源：lens.org

1.3 高质量研究往往表现出延迟承认特征

上文探讨了睡美人文献的技术与应用属性，特别是与专利引用的相关性，接下来再探讨睡美人文献与学术质量之间的相关性。

1.3.1 学术质量与学术影响之差异的相关研究

学术质量和学术影响之间的相关性研究或两者之间差异之研究一直是科学计量学与科技评价领域的一个热点研究主题。目前，以诺贝尔奖论文作为高质量论文代表探讨被引次数与学术质量的相关研究主要有两项：一是美国 Eagle 分子医学研究所的 Shi V. Liu 统计了 2002-2005 年诺贝尔生理学或医学奖、化学奖、物理学奖、经济学奖的代表性论文的被引次数。代表性论文通过 2 种方式确定：①诺奖得主本人在诺奖颁奖典礼或个人简历上的权威描述；②诺奖委员会在当年的诺奖发布新闻稿上提供的信息。研究指出包括诺贝尔奖在内的许多重要突破性工作在被引频次方面表现一般或低引用[12]。二是我国学者叶鹰和李江的相关研究。叶鹰采用 20 世纪 Nature 上发表的 21 项被认为改变了科学和世界的重要发现[13]作为高质量论文代表，发现高品质论文不是最高被引论文，但都高于学科平均被引[14]。李江和叶鹰继而采用对照研究发现，高质量论文的被引曲线多数不规则，而高影响论文的被引曲线多数很规则[15]。根据上述启发，笔者仍以诺贝尔奖关键论文为对象，对其被引延迟程度或睡美人文献特征进行测度和研究。考虑到最近 10~20 年科学技术的发展速度、学术共同体和科技论文总量规模的膨胀，选取最近授予的诺贝尔奖及其关键论文为对象展开分析。

1.3.2 诺奖关键论文是否遭遇延迟承认？

（1）资料与方法

为了寻找获得诺奖的关键文献（key Publications），笔者系统研究了诺奖官网（http://www.nobelprize.org/）的相关信息[16]，发现 2009-2016 年的诺奖官网专门标注了生理学或医学奖获奖的代表作，在 Press Release 或 Scientific Background 中以加粗、加星号或专门注明的形式给出。除 2014 年的化学奖以外，其他领域各年获奖均未给出代表作。获取到 2009-2016 年度授予的诺贝尔生理学或医学奖的关键论文共计 33 篇，通过 Web of Science 统计被引次数，截至 2016 年 11 月 18 日。

笔者引入 Wang(2013)提出的被引速率（Citation Speed, CS）指标，即整体上一篇论文自发表后以多快的速度累积它的被引次数[17]。对于快速突破型论文，被引次数会在发表后

迅速累积，到达一个较高的水平，接下来维持稳定，整体上的被引速率很快（高）。但对于延迟承认型（睡美人）文献，发表之初的被引次数累积得很少、很慢，直到最后几年才累积完，所以整体上的被引速率较低。如果一篇论文以恒定的速率累积其被引次数，说明每年的被引次数相同。Wang, Thijs& Glänzel (2015)提出了被引速率的反向指标——Citation Delay，简称 D。笔者将其翻译为被引延迟指数[18]，其定义是：

$$\text{Citation Delay}=1-\text{Citation Speed}$$

D 反映了论文的被引延迟程度，D 值越大，说明论文被引的延迟程度越大。我们采用 D 指标反映在观察期内论文总被引次数累积的快慢。被引延迟指数大于 0.5，说明整体上被引次数延迟，小于 0.5 说明整体上论文被引次数是在前半段累积起来的。设置两个观察期，一是截至本研究的时间；二是截至获得诺贝奖的时间。

（2）三分之二的论文遭遇延迟承认

33 篇诺奖关键论文信息见下表，Nature、Cell 和 Science 仍是发表诺奖成果论文的主要期刊，分别发表 9 篇、6 篇和 3 篇，占 54.5%，即超过一半的医学领域诺奖关键论文发表在这三大权威期刊上。33 篇论文，最早发表于 1962 年，最近发表于 2006 年。

表 2 2009-2016 年授予的诺贝尔生理学或医学奖的 33 篇关键论文

获奖年	获奖者	序号	期刊	被引次数	发表年	D1	D2
2016	Yoshinori Ohsumi	1	Journal of cell biology	564	1992	0.590	0.590
		2	Febs letters	733	1993	0.663	0.663
		3	Nature	733	1998	0.607	0.607
		4	Nature	819	2000	0.620	0.620
2015	Satoshi Ōmura	5	Antimicrobial agents and chemotherapy	588	1979	0.535	0.988
	William C. Campbell	6	Antimicrobial agents and chemotherapy	302	1979	0.382	0.987
	Youyou Tu	7	Yao xue xue bao	14	1981	0.761	0.995
2014	John O'Keefe	8	Brain research	2150	1971	0.781	0.781
		9	Experimental neurology	747	1976	0.647	0.638
	May-Britt Moser and Edvard I. Moser	10	Science	495	2004	0.582	0.603
		11	Nature	1074	2005	0.632	0.644
		12	Science	466	2006	0.598	0.613
2013	James E. Rothman	13	Cell	562	1984	0.393	0.416
	Randy W. Schekman	14	PNAS	364	1979	0.508	0.522
		15	Cell	549	1990	0.367	0.387
	Thomas C. Südhof	16	Nature	638	1990	0.385	0.410
		17	Nature	2331	1993	0.371	0.387
		18	Nature	514	1993	0.421	0.443
2012	Sir John B. Gurdon	19	Journal of embryology and experimental morphology	435	1962	0.732	0.649
	Shinya Yamanaka	20	Cell	9271	2006	0.634	0.720
2011	Bruce A. Beutler	21	Science	4964	1998	0.506	0.552
	Jules A. Hoffmann	22	Cell	2136	1996	0.557	0.595
	Ralph M. Steinman	23	Journal of experimental medicine	1320	1973	0.647	0.612
		24	PNAS	624	1978	0.423	0.408

		25	Journal of experimental medicine	948	1985	0.407	0.450
2010	Robert G. Edwards	26	Lancet	200	1965	0.558	0.537
		27	Nature	277	1969	0.457	0.424
		28	Nature	128	1970	0.385	0.338
		29	Lancet	905	1978	0.674	0.651
		30	Nature medicine	14	2001	0.681	0.889
2009	Jack W. Szostak; Elizabeth H. Blackburn	31	Cell	391	1982	0.378	0.332
	Carol W. Greider; Elizabeth H. Blackburn	32	Cell	1960	1985	0.622	0.645
	Carol W. Greider; Elizabeth H. Blackburn	33	Nature	1062	1989	0.501	0.519

注：D1 是指从论文发表年到被引次数统计年的被引延迟指数；D2 是指从论文发表年到获奖年的被引延迟指数。

对 33 篇文献的被引延迟程度测度结果表明，无论是被引次数的统计时间界定为至获奖年还是至 2016 年，三分之二的文献（22 篇）的被引延迟指数均大于 0.5，说明从长期来看，这些文献都遭遇了一定程度的延迟承认，这对延长科研评价周期有重要启示。

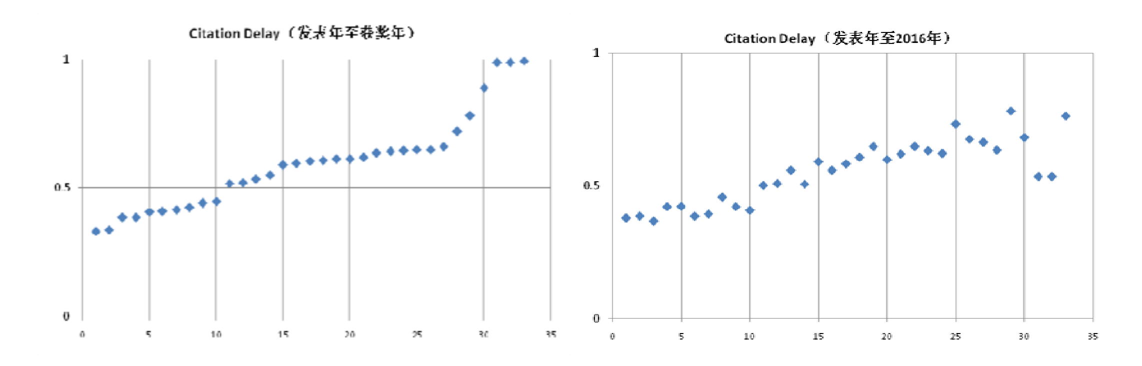


图 2 2009-2016 年授予的诺贝尔生理学或医学奖的 33 篇关键论文的被引延迟指数

上文归纳了睡美人文献的 3 个重要特征，跨学科性，技术与应用属性，多为高质量研究。基于上述研究，笔者试图凝练出预测睡美人文献的一些关键线索和思路。

2 睡美人文献的预测线索

睡美人文献的预测属于科技预测的范畴，任何预测都很困难。对于睡美人文献，要想检验预测是否准确，还需要长期的观察，需要时间因素。本文并未直接提出睡美人文献的预测方法或综合模型，而是通过基础性研究，希望能为预测睡美人文献提供一些关键的线索或思路。主要体现在两个方面：

2.1 识别变革性研究并追踪其技术转化应用状况

识别变革性研究并追踪研究转化或被专利引用情况可作为预测睡美人文献的重要线索。无论是通过典型睡美人文献案例的回顾性研究，还是笔者前期基于专家评议信息追踪论文被引次数状况的前瞻性研究[19]都提示，睡美人文献反映的创新成果多是变革性研究，实际上，变革性研究是睡美人文献这一现象的本质。从 Hell(1994)案例来看，预测睡美人文献，要特别关注变革性研究，特别是那些提出可以打破某一领域经典范式的新方法或新观点的文献——在标题或摘要中常出现“break/overcome…limits”、“possible”、“potential”等术语。对于

这类文献的作者，要注意追踪检索其后续发表的文献，如果发现他/她仍然在坚持该领域的研究，而且一段时间之后能够有成功的表现（例如，提出的理论方法在实际应用中获得成功，且研究成果发表于高影响力期刊），或可以大胆预测，作者提出初始思路的那篇文章有可能就是睡美人文献。

另外，本文又在 van Raan 的研究基础上，验证了睡美人文献潜在的技术与应用属性，而非人们通常所认为的睡美人文献应该多为高深莫测的理论性、基础性研究的传统认识。从研究层次（Research level）上看，睡美人文献的研究内容多为技术创新或应用研究，早期即被专利引用。所以，如果一篇论文反映的创新内容属于变革性研究，比如在标题或摘要中明显标注类似术语（如 disagree、overcome、break、dispute…等），如果这样的研究论文发表后多年一直未被引用或被引很少，但却发表之后立即被专利引用，说明该研究具有技术转化潜力，论文沉睡可能是因为技术尚未成熟而未被大量引用，而一旦后期条件或技术成熟，或许能带动那篇文章的被引次数增长。所以，要预测睡美人文献，首先要识别变革性研究，然后做好"追踪"，找到如下蛛丝马迹：比如作者是否持续开展该主题的研究，是否从理论研究深化、拓展到实践研究，实践中是否成功；论文发表之后是否有专利授权，论文是否被专利引用等，这些都是预测零被引或低被引论文在未来获得高被引潜力的重要线索。

2.2 应用 Bcp 指数识别正处于"沉睡-唤醒"萌芽期的研究

笔者在 Ke et al.(2015)提出的美丽指数（*B* 指数）基础上，通过将考察对象"年度被引次数"优化为"年度被引次数的累积百分比"，提出一个用于识别睡美人文献的新的简易无参数指标——*Bcp* 指数，并在此框架下重新定义了唤醒时间、睡眠时长和睡眠深度[20]。实证研究发现，*Bcp* 指数对论文发表之初被引次数的约束效力显著高于 B 指数，更符合睡美人文献早期零被引和低被引的特征。*Bcp* 指数对论文总被引次数和年度被引次数最大值的依赖程度低于 B 指数。我们关注零被引或低被引现象的意义在于发现低被引文献中意义重大，即成为睡美人的文献。*B* 指数对总被引次数的依赖性高，说明关注点仍侧重（最）高被引文献，可能忽视了次高被引文献中睡美人的存在，自然也忽视了剖析其中的规律。在文献计量学中，最高被引文献通过被引次数这一单一指标即可识别出来，不易被忽视。而把被引次数相对不是最高、平时不易观察到，但却又具有睡美人特征、有潜在重大意义的文献识别出来，具有重要意义。*Bcp* 指数对早期发表的睡美人文献更敏感，且更容易识别出总被引次数和年度被引次数较低的睡美人文献。

笔者认为，Bcp 指数可作为预测睡美人文献的另一思路。笔者已通过数学公式推导出，对于任意一篇非零被引论文，Bcp 指数的取值范围是[-(a-1)/2, (a-1)/2]，a 为论文年龄（age）；Bcp=0，这说明年度被引次数相同。Bcp 指数的优势在于考察的是论文年度被引次数累积百分比，所以对于不同的被引次数规模的论文的延迟承认程度均可以进行直接比较。例如对于一篇年龄为 10 年的论文，Bcp 值约接近于 4.5，则说明延迟承认程度越高。Bcp 指数的优势就在于即使一篇论文从发表至考察期被引次数仍较少，但能够识别出那些正处于"沉睡-唤醒"萌芽期的论文，比如首次出现被引突增苗头的论文。有被引突增的苗头，则可能预示着未来这种增长势头会继续显现，而成为睡美人文献。关于采用一个综合模型来预测睡美人文献，是一个未来努力的方向或目标。

3 睡美人文献研究的科技政策启示

基于上文分析，本部分从加强变革性研究、识别研究前沿和改善学术评价三个方面讨论分析挖掘睡美人文献的科技政策启示与建议。

3.1 睡美人文献与变革性研究

睡美人文献反映的创新成果多是变革性研究，这类研究若遭遇延迟承认，可能存在跨领域唤醒机制，即在一个领域提出的创新思想可能在新的领域有了用武之地。变革性研究是指

挑战或颠覆既有研究范式的创新性研究。这一概念主要来源于美国国家科学基金会（NSF），也受到了我国科技政策制定者和政策分析专家的密切关注[21],[22]。《"十三五"国家科技创新规划》提出"推进颠覆性技术创新，加速引领产业变革"，实现颠覆性技术创新离不开变革性研究；《国家自然科学基金"十三五"发展规划》也提到，完善从非共识学术争议中甄别发现原创思想的工作机制，探索加强对挑战传统范式的变革性研究的资助力度。但由于变革性研究挑战或颠覆现有研究范式，让同行专家取得共识性的评议意见非常困难。如何使得同行评议专家积极发现和及时甄别具有变革性创新的项目申请书，仍然面临着诸多挑战。随着交叉学科的不断兴起和发展，如何做好交叉学科科研项目的评定和资助工作，如何早期识别和评价出某项颠覆性创新研究的价值，都还需要基金资助机构进一步探索创新同行评议方式。

3.2 睡美人文献与研究前沿

根据创新程度，科研成果可分为开创性成果与扩展性成果。睡美人文献与研究前沿密切相关。与热点跟风式成果相比，重大创新成果往往遭遇延迟承认，这是科技创新的基本规律。科学界有这样一种共识——"科研做的年头一长，我们就会渐渐明白，论文的影响和学术水平其实不是一回事。好的科研往往独辟蹊径，所出的成果需要过一段时间甚至很多年以后才会慢慢被主流接受，在短期引用率上反映不出来。论文 IF 的计算基于短期（两年）引用，所以要追求论文的 IF，意味着我们必须做热点性、跟风性的研究工作。"[23]受上述共识启发，笔者认为可以用延迟承认的思路开展研究前沿的识别。情报学擅长的研究热点分析多以"快速高被引"的论文为基础资料，而科学家眼中的研究前沿则可能是缓慢高被引、目前处于爆发前萌芽状态、引文曲线符合睡美人文献特征（一开始被引很少，后来突然高被引）的研究内容。因此，作为文献计量学特定概念的研究前沿（research front）和科学家眼中的研究前沿（research frontier）不完全等同。

高被引论文可分为"最高被引论文"和"次高被引论文"。关注低被引现象的意义在于发掘低被引论文中意义重大的论文。情报学领域擅长做"研究热点"，但如何分析"研究冷门"，特别是要识别出有潜力、重要的研究冷门，仍是一个挑战。根据科学发展的自身规律，真正好的、创新性的研究总是少数人开创的、容易被忽视或抵制、开始阶段不可能纷纷跟风涌入的。Bcp 指数可用于挖掘正处于沉睡-唤醒萌芽期的论文。利用 Bcp 指标，能够识别出总被引次数不是那么高，但生命周期的后段被引次数突增的这类文献，能够找到一些科学界里面处在沉睡-唤醒这个导数特别大的区间的工作，进而推动科学界及早关注这一类研究。把这类平时不易观察到，但又有潜在重大意义的研究工作识别出来，对于前瞻部署科技前沿、识别有潜力的研究冷门、乃至预测诺贝尔奖得主等提供了新的思路、工具和方法。

3.3 睡美人文献与学术评价

当前政府和科技界正在倡导建立以科技创新质量、贡献、绩效为导向的分类评价体系。科技评价应侧重强调研究的质量和贡献。一篇论文的影响力取决于其所在的学科规模、论文主题的研究热度、理论贡献、作者声誉、期刊影响力等诸多因素，而一篇论文的质量主要取决于理论贡献。以往研究和本研究均表明，高质量论文不一定高被引，高品质、重大原创性论文往往会有一个"沉睡期"，会遭遇延迟承认，在被引次数上表现出睡美人文献的特征。因此，本文所提出的 Bcp 指数可作为评价论文学术质量一个参考指标。

科技论文发表之后要经受形式评价、内容评价和历史评价。第一种评价依据刊物影响因子、领域排序、论文引用等指标；第二种依据同行对研究者所报告信息的分析比较；第三种则依据经受了时间检验的科学贡献[24]。鉴于重大科技创新成果与学术热点跟风式成果相比往往遭遇延迟承认，目前学术评价实践中大都将"近 5 年代表性论文及其被引次数"作为一个必要指标，建议遵循科技创新的基本规律，探索建立长效评价机制，适度延长评价周期。

参考文献

[1] van Raan AFJ. Sleeping beauties in science[J]. Scientometrics, 2004, 59(3): 467–472.

[2] Palomeras N. Sleeping patents: any reason to wake up?[J]. Iese Research Papers, 2003, 20(35):1-37.

[3] 董尔丹，胡海，张俊．学术评价应更科学[J]．科学通报, 2014, 59(1):96-106.

[4] Wang J, Ma F, Chen M, et al. Why and how can "sleeping beauties" be awakened?[J]. The Electronic Library, 2012, 30(1):5-18.

[5] van Raan A F J. Dormitory of physical and engineering sciences: Sleeping beauties may be sleeping innovations[J]. PloS one, 2015, 10(10): e0139786.

[6] van Raan A F J. Sleeping Beauties Cited in Patents: Is there also a Dormitory of Inventions?[J]. 2017, 110(3): 1123-1156

[7] 杜建，武夷山．基于被引速率指标识别睡美人文献及其"王子"——以 2014 年诺贝尔化学奖得主 Stefan Hell 的睡美人文献为例．情报学报, 2015,34(5): 508-521.

[8] Hell S W, Wichmann J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy[J]. Optics Letters, 1994, 19(11):780-782.

[9] Hell S W, Kroug M. Ground-state-depletion fluorscence microscopy: A concept for breaking the diffraction resolution limit[J]. Applied Physics B, 1995, 60(5):495-497.

[10] Du J, Wu Y. A Bibliometric Framework for Identifying "Princes" Who Wake up the "Sleeping Beauty" in Challenge-type Scientific Discoveries[J]. Journal of Data & Information Science, 2016(1):50-68.

11杜建，武夷山．一个识别睡美人文献的新的无参数指标——基于 Science 和 Nature 上睡美人文献的验证．情报理论与实践，2017,40（2）：19-25

[12] Shi V. Liu. Nobel Prize-Winning Original Publications' Under Performance in Making Citation Glory. Logical Biology. 2005, 5(4): 297-305

[13] Garwin, Laura, and Tim Lincoln, eds. A century of Nature: twenty-one discoveries that changed science and the world. University of Chicago Press, 2003.

[14] 叶鹰.高品质论文被引数据及其对学术评价的启示[J].中国图书馆学报,2010,36(1):101-104

[15] Li J, Ye F Y. A probe into the citation patterns of high-quality and high-impact publications[J]. Malaysian Journal of Library & Information Science, 2014, 19(2):17-33.

[16] All Nobel Prizes: http://www.nobelprize.org/nobel_prizes/lists/all/index.html

[17] Wang J. Citation time window choice for research impact evaluation[J]. Scientometrics, 2013, 94(3): 851-872.

[18] Wang J, Thijs B, Glänzel W. Interdisciplinarity and impact: distinct effects of variety, balance, and disparity[J]. Plos One, 2015, 10(5):e0127298.

[19] Du J, Tang X, Wu Y. The effects of research level and article type on the differences between citation metrics and F1000 recommendations[J]. Journal of the Association for Information Science & Technology, 2016, 67(12):3008-3021.

[20] 杜建，武夷山．一个识别睡美人文献的新的无参数指标——基于 Science 和 Nature 上睡美人文献的验证．情报理论与实践，2017,40（2）：19-25

[21] 龚旭．科学基金与创新性研究——美国国家科学基金会支持变革性研究的相关政策分析[J]．中国科学基金, 2011(2):105-110.

[22] 郑永和，陈淮．美国国家科学基金会加强支持变革性研究考察[J]．中国基础科学, 2008, 10(4):39-42.

[23] 贾伟. IF 决定命运[EB/OL]. (2017-02-13).[2017-08-14]. http://blog.sciencenet.cn/blog-265898-1033254.html

[24] 徐旭东．中国 CNS 论文与科学贡献评价[EB/OL]. (20150608) [2017-07-23]. http://blog.sciencenet.cn/blog-1636782-896460.html